Regular Paper

Ameliorating Mental Mistakes in Tradeoff Studies*

Eric D. Smith,[1,2,†] Young Jun Son,[2] Massimo Piattelli-Palmarini,[3] and A. Terry Bahill[2,‡]

[1]*Engineering Management and Systems Engineering, University of Missouri-Rolla, Rolla,* MO 65409-0370

[2]*Department of Systems and Industrial Engineering, University of Arizona, Tucson,* AZ 85721-0020

[3]*Department of Psychology, University of Arizona, Tucson,* AZ 85721-0068

Received 26 June 2006; Revised 18 December 2006; Accepted 7 March 2007, after one or more revisions
Published online in Wiley InterScience (www.interscience.wiley.com).
DOI 10.1002/sys.20072

ABSTRACT

Tradeoff studies are broadly recognized and mandated as *the* method for simultaneously considering multiple alternatives with many criteria, and as such are recommended in the Capability Maturity Model Integration (CMMI) Decision Analysis and Resolution (DAR) process. Tradeoff studies, which involve human numerical judgment, calibration, and data updating, are often approached with under confidence by analysts and are often distrusted by decision makers. The decision-making fields of Judgment and Decision Making, Cognitive Science and Experimental Economics have built up a large body of research on human biases and errors in considering numerical and criteria-based choices. Relationships between experiments in these fields and the elements of tradeoff studies show that tradeoff studies are susceptible to human mental mistakes: This paper indicates ways to eliminate the presence, or ameliorate the effects of mental mistakes on tradeoff studies. © 2007 Wiley Periodicals, Inc. Syst Eng 10: 222–240, 2007

Key words: tradeoff studies; trade studies; cognitive biases; decision analysis; problem statement; evaluation criteria; weights of importance; alternative solutions; evaluation data; scoring functions; scores; combining functions; preferred alternatives; sensitivity analysis; decision making under uncertainty

---

*This work was supported by the Air Force Office of Scientific Research under AFOSR/MURI F49620-03-1-0377.
† Present address: Department of Engineering Management and Systems Engineering, University of Missouri at Rolla, MO 65409.
‡Author to whom all correspondence should be addressed (e-mail: terry@sie.arizona.edu; son@sie.arizona.edu; Massimo@u.arizona.eduerics@umr.edu).

1. INTRODUCTION

A systems engineer may be tasked with capturing the values and preferences of a decision-making customer, so that the decision-maker and other stakeholders will have confidence in their decisions. However, biases, cognitive illusions, emotions, fallacies, and the use of simplifying heuristics make humans far from ideal decision-makers and makes capturing their preferences challenging. Using tradeoff studies judiciously can help people make rational decisions. But they have to be careful to make sure that their tradeoff studies are not plagued with mental mistakes. This paper presents lessons learned in ameliorating mental mistakes in doing tradeoff studies.

Tradeoff studies, which are often called trade studies, provide an ideal, rational method for choosing among alternatives. Tradeoff studies involve a mathematical consideration of many evaluation criteria for many alternatives simultaneously, in parallel. Without a tradeoff study, short-term attention usually leads people to consider criteria serially. Benjamin Franklin, cited in MacCrimmon [1973] and Bahill [2007], wrote that when people draw out the consideration of criteria over several days, thoughts focused on different criteria naturally arise separately in time. Tversky and Kahneman [1974, 1981] showed that humans easily anchor on certain values before evaluating probabilities [1974] or numbers [1981], thus causing mistakes. Such phenomena seem robust, as recent research has shown that different areas of the brain are involved in different types of decisions, and that framing and bias show neurobiological correlates [Martino De et al., 2006]. Framing of hypotheses is a general problem for human decision making [Petroski, 2003]. Fixation on short-term rewards is also a well-documented problem in the absence of a tradeoff study [Ainslie, 2001].

Tradeoff studies are probably as old as mathematics. The basic procedure of a tradeoff study is to account numerically for empirical values and preferences. When this accounting is formalized, it provides a stable base for deciding among alternatives. Aristotle (384–322 BC) developed logic and empirical observation, and noted that humans are capable of deliberating on a choice among alternatives. An early description of a tradeoff study that summed weighted scores is given in a 1772 letter from Benjamin Franklin to Joseph Priestley [MacCrimmon, 1973] and [Bahill, 2007].

The seminal academic descriptions of tradeoff studies appeared in Keeney and Raiffa [1976] and Edwards [1977]. Edwards focused on the numerical determination of subjective values, an activity that has been shown to be difficult by the cognitive sciences. Keeney and Raiffa are best known for their axiomatic derivation of value and utility functions from conditions of preferential or utility independence. In practice, it is expensive and difficult to design and implement experiments that demonstrate these independence conditions, or measure input values among the criteria (attributes) of alternatives. Moreover, searches for alternatives usually produce complex alternatives that are so different from one another that it is difficult to invoke these elegant mathematical theorems. Haimes [2004: 208] notes that some discussions of optimality conditions for tradeoff approaches are "strictly theoretical, however, since it is difficult or even impossible to implement the utility function approach in practice." This is so because "it is extremely difficult or impossible to actually determine the decision maker's utility function—that is, to assign numerical utilities to the various combinations of the objectives" [Haimes, 2004: 209].

The INCOSE Systems Engineering Manual [INCOSE, 2004] provides a consensus of the fundamentals of tradeoff studies. Tradeoff studies are prescribed in industry for choosing and ranking alternative concepts, designs, processes, hardware, and techniques. Today, tradeoff studies are broadly recognized and mandated as *the* method for choosing alternatives by considering many criteria simultaneously. They are the primary method for choosing among alternatives listed in the Software Engineering Institute's Capability Maturity Model Integration [CMMI, 2006] Decision Analysis and Resolution [DAR, 2004] process. This paper seeks to make tradeoff studies more robust.

This paper suggests that even if you do all of the mathematics correctly and furthermore do all of the utility curves correctly, you still have to be careful in doing a tradeoff study, because it is hard to overcome mental mistakes. This paper point outs such mental mistakes and offers some suggestions for ameliorating them. Individual systems engineers can use this knowledge to sharpen their own decision processes and institutions can incorporate this knowledge into their documented decision-making processes.

2. COMPONENTS OF A TRADEOFF STUDY

This section describes the components of a tradeoff study, including the (1) problem statement, (2) evaluation criteria, (3) weights of importance, (4) alternate solutions, (5) evaluation data, (6) scoring functions, (7) scores, (8) combining functions, (9) preferred alternatives, and (10) sensitivity analysis.

Problem statement. Stating the problem properly is one of the systems engineer's most important tasks, because an elegant solution to the wrong problem is less than worthless. Problem stating is more important than problem solving [Wymore, 1993: Section 1.1]. The problem statement describes the customer's needs, states the goals of the project, delineates the scope of the system, reports the concept of operations, describes the stakeholders, lists the deliverables, and presents the key decisions that must be made. The problem must be stated in a clear, unambiguous manner. "The problem of the design of a system must be stated strictly in terms of its requirements, not in terms of a solution or a class of solutions"—Wayne Wymore [2004, p. 9].

Evaluation criteria are derived from high-priority tradeoff requirements. Each alternative will be given a value that indicates the degree to which it satisfies each criterion. This should help distinguish between alternatives. Criteria should be independent, but they should also show compensation, which means that they can be traded off: For example, we might prefer a car with less horsepower, if it is less expensive. Criteria must be arranged hierarchically: the top-level categories might be performance, cost, schedule, and risk. Each company should have a repository of generic evaluation criteria that are tailored for each specific project. Evaluation criteria are called measures of effectiveness by SysML [http://www.sysml.org].

Weights of importance. The decision analyst should assign weights to the criteria so that the more important ones will have more effect on the outcome. Weights are often given as numbers between 0 and 10, but are usually normalized so that in each category they sum to 1.0. There are a dozen methods for deriving numerical values for the weights of importance for the evaluation criteria [Buede, 2000; Daniels, Werner, and Bahill, 2001; Kirkwood, 1999; Weber and Borcherding, 1993]. These methods can be used by individuals or teams. If pairwise comparisons of preferences between the criteria can be elicited from experts, then the weights of importance can be determined through the Analytic Hierarchy Process (AHP) [Saaty, 1980: 20–21]. However, as with most systems engineering tools, AHP has both loathers and zealots. Weights can also be derived with the method of swing weights, wherein weights of importance are derived by swinging a criterion from its worst value to its best value [Keeney and Raiffa, 1976]. In the case of tradeoff studies at many levels, the customer or stakeholders should establish weights for key criteria during the conceptual design. These weights should then flow down to tradeoff studies in the detailed design [Wymore, 1993; Buede, 2000].

Alternative solutions must be proposed and evaluated. In addition to plausible alternatives, the do nothing alternative (often the status quo) and alternatives with extreme criteria values must be included to help get the requirements right and to test the structure of the tradeoff study. If the do nothing alternative or another alternative with an extreme criteria value (or values) wins the tradeoff analysis, then it is probable that either the requirements or the structure of the tradeoff study is wrong.

Evaluation data come from approximations, product literature, models, analyses, simulations, experiments, and prototypes. Evaluation data are measured in natural units, and indicate the degree to which each alternative satisfies each criterion. The collection, preparation, and choice of the evaluation data can be subject to subtle biases and errors, especially if the data are subjective or elicited. Section 3 discusses cognitive impediments to tradeoff studies, preference biases, probabilistic fallacies, and circumstances that affect the collection and selection of quantitative evaluation data. Evaluation data could, of course, have measures of variance attached. These measures of variability could be propagated through the whole tradeoff study and be attached to the preferred alternatives; but this is seldom done in practice.

Scoring functions. Evaluation data are transformed into normalized scores by scoring functions (utility curves) or qualitative scales (fuzzy sets) [Wymore, 1993: 385–398; Daniels, Werner, and Bahill, 2001]. The shape of scoring functions should ideally be determined objectively, but often subjective expert opinion is involved in their preparation. Creating scoring function packages takes a lot of time. A scoring function package should be created by a team of engineers, and be reevaluated with the customer with each use.

Scores. The numerically normalized 0 to 1 scores obtained from the criteria scoring functions are easy to work with. Assuming that the weights of importance are also normalized, combining these scores leads to a rankable set of alternative ratings that preserve the normalized 0 to 1 range.

Combining functions. The weights and scores must be combined in order to select the preferred alternatives. The most common combining functions are

$$\text{Sum Combining Function} = wt_x x + wt_y y,$$

$$\text{Product Combining Function} = x^{wt_x} \times y^{wt_y},$$

$$\text{Sum Minus Product Combining Function} = wt_x x + wt_y y - x^{wt_x} \times y^{wt_y},$$

$$\text{Compromise Combining Function} = [(wt_x x)^p + (wt_y y)^p]^{1/p},$$

where x and y are the values of the criteria and wt_x and wt_y are the weights of importance. Combining functions are called objective functions by SysML, although the optimization and control theory communities use the phrase objective functions differently. One must be careful to choose a combining function appropriate to the situation, and not be a victim of an odd mathematical vagary of a combining function [Smith, 2006]. In addition to these four, dozens of other combining functions have been proposed in the literature; each has been "proven" to be most appropriate for specific situations arising from specific axioms or assumptions [Keeney and Raiffa, 1976]. The Sum Combining Function is used in most implementations of quality function deployment (QFD), which is often called the House of Quality [Bahill and Chapman, 1993]. QFD is useful for evaluating and prioritizing criteria and alternatives.

Preferred alternatives should arise from the impartial, parallel consideration of the weights of importance and the evaluation criteria scores for each alternative. Each alternative's final rating will allow a ranking of alternatives. Care must be taken, however, to eliminate human tendencies that may draw the study to a result that is politically preferred.

Sensitivity analysis identifies the most important parameters in a tradeoff study; often these are the cost drivers that are worthy of further investment. A sensitivity analysis of the tradeoff study is imperative. In a sensitivity analysis, you change a parameter or an input, and measure changes in outputs or performance indices [Karnavas, Sanchez, and Bahill, 1993]. You should perform a sensitivity analysis anytime you perform a tradeoff study, create a model, write a set of requirements, design a system, make a decision, or search for cost drivers. A sensitivity analysis completes a tradeoff study by gauging the robustness of the preferred alternatives and the rationality of the overall study.

Tradeoff study components. Evaluation Criteria are derived from a Problem Statement, and possible Alternatives are selected. In a preliminary screening stage, mandatory requirements or screening constraints are used to reduce the number of alternatives that must be considered. In this stage, sometimes referred to as "elimination by aspects" or "killer trades," alternatives that do not meet the screening constraints are dropped. In the computational secondary evaluation stage, Evaluation Data for each Evaluation Criterion are normalized with a Scoring Function, and are combined according to the Weights of Importance and the Combining Functions, yielding a final Rating for each alternative. A Sensitivity Analysis is conducted to determine the robustness, and a list of Preferred Alternatives is written. The formal tradeoff study report is peer-reviewed, and the results are given to the originating decision-maker and are put into a process asset library (PAL). Of course, as with all systems engineering processes, tradeoff studies are not performed in a serial manner. Tradeoff studies must be performed iteratively, with many parallel feedback loops.

Tradeoff studies are performed at the beginning of a project to help select the desired system architecture and make major purchases. However, throughout a project you are continually making tradeoffs: creating team communication methods, selecting tools and vendors, selecting components, choosing implementation techniques, designing test programs, and maintaining schedule [Bahill and Briggs, 2001]. Many of these tradeoffs should be formally documented.

"Hence, there are two requirements for assessing good value trade-offs. The first is to do the right thing (i.e. focus on the logical substance of the value trade-off problems), and the second is to do it right (i.e. avoid the psychological traps that influence assessment procedures)" [Keeney, 2002: 943]. Doing the right thing means having the will to look at the problem logically. Of the ten tradeoff study components presented above, the Problem Statement deals with selecting the right problem or doing the right tradeoff study, while the next sections of this paper give recommendations for doing a tradeoff study right.

3. MENTAL MISTAKES THAT CAN AFFECT COMPONENTS OF TRADEOFF STUDIES

Study of the tradeoff process can help explain why the detailed completeness of tradeoff studies are often deemed unnecessary by overly confident decision-makers, and can help eliminate biases from the mechanics of the tradeoff process. Many conclusions obtained from Judgment and Decision Making, Cognitive Science and Experimental Economics can be used to shed light on various aspects of the tradeoff process. Of course, since many experiments in these fields were designed to reveal truths about human choice at a basic level, they do not exactly model the processes of tradeoff studies. Therefore, in the following sections, the elements of the experiments and the components of tradeoff studies may be compared on an abstract level.

This decision-making literature uses the terms biases, cognitive illusions, emotions, fallacies, attribute substitution, simplifying heuristics, fear of regret, psychological traps, and paradoxes. In this paper, when we refer to specific experiments we will use the term used in the basic research papers. However, when we show how these things could adversely affect a tradeoff study, we will collectively call them mental mistakes.

Humans often make mental mistakes in conducting tradeoff studies. Smith [2006] extracted seven dozen heuristics and biases (including representiveness, anchoring, base-rate neglect, and the conjunction fallacy), cognitive illusions, emotions, fallacies, fear of regret, psychological traps and paradoxes from the psychology, decision-making, and experimental economics literature and showed how they could induce mental mistakes in tradeoff studies. A matrix of relations between cognitive biases and tradeoff study components is available at http://rayser.sie.arizona.edu:8080/resume/Seminar/MatrixOfBiases.zip. Many of these are mentioned in Piattelli-Palmarini [1994], Beach and Connolly [2005], Hammond, Keeney, and Raiffa [2006], Shafir [2004], and the seminal work of Kahneman, Slovic, and Tversky [1982]. Sage [1990] covered different models of analytic and intuitive thinking early on in the systems engineering literature.

We studied these cognitive biases, and determined whether each bias principally affects (1) magnitudes, such as evaluation data, weights, and probabilities, or (2) text, such as stating the problem, writing criteria or proposing alternatives. We then categorized the biases by tradeoff study component.

For this paper, we also used tradeoff studies collected by Bahill over the last two dozen years from a Systems Engineering Process class at the University of Arizona and from industry files observed by Bahill during summer industry work and academic sabbaticals. In the systems engineering course, students in teams of three or four wrote the Wymorian eight-document sets [Wymore, 1993] for a particular system design problem. On average, each set of documents comprised 100 pages and took at least 100 person hours to write. One of each team's documents contained a system-level tradeoff study. Some of these document sets are available at http://www.sie.arizona.edu/sysengr/sie554/. We studied about 200 of these tradeoff studies looking for mental mistakes. The mental mistakes that we found are summarized in this paper.

The next section describes these mental mistakes and gives recommendations for ameliorating their effects on the tradeoff study components. The present work draws on existing, rigorous experimental results in the cognitive science literature and serves as a warning of the large-scale possible occurrence of cognitive biases within tradeoff studies. This work thus submits guidelines for consideration within the systems engineering community. The positive experimental identification of each bias—specifically within tradeoff studies and tradeoff study components—is a goal generally not pursued, so experimentation on all the biases in a tradeoff study context could take decades.

A Brief Discussion of Prospect Theory

The seminal paper on cognitive biases and heuristics under uncertainty is by Tversky and Kahneman [1974]; this work engendered Prospect Theory [Kahneman and Tversky, 1979]. Prospect Theory breaks subjective decision-making into a preliminary screening stage and a secondary evaluation stage. In the screening stage, values are considered not in an absolute sense (from zero), but subjectively from a reference point established by the subject's perspective and wealth before the decision. In this stage, losses are weighted more heavily than gains. In the evaluation stage, a final value for every prospect (alternative) is calculated. Later refinements to this theory, including subjective probability assessment, were published in Tversky and Kahneman [1992], which was reprinted in Tversky and Kahneman [2000]. Kahneman won the Nobel Prize in Economics in 2002 "for having integrated insights from psychological research into economic science, especially concerning human judgment and decision-making under uncertainty" [RSAS, 2002] [Kahneman, 2003].[1]

Several heuristics and biases, notably representiveness, anchoring, base-rate neglect, and the conjunction fallacy are now considered by Kahneman to be instances of a super-heuristic called *attribute substitution.* Judgment is mediated by this heuristic when, without realizing that it is so,

> an individual assesses a specified target attribute of a judgment object by substituting another property of that object—the heuristic attribute—which comes more readily to mind. Many judgments are made by this process of *attribute substitution.* For an example, consider the well-known study by Strack, Martin, & Schwarz [1988], in which college students answered a survey that included these two questions: How happy are you with your life in general? How many dates did you have last month? The correlation between the two questions was negligible when they occurred in the order shown, but it rose to 0.66 when the dating question was asked first" [Kahneman and Frederick, 2002: 53].

As we see clearly in this instance, the target attribute (happiness) is assessed by mapping the value of another attribute (number of dates last month) onto the target scale. This process of attribute substitution

> will control judgment when these three conditions are satisfied: (1) the target attribute is relatively inaccessible (happiness); (2) a semantically and associatively related candidate attribute is highly accessible (number

[1]Tversky did not share this prize because Nobel Prizes are not bestowed posthumously.

> of dates last month); and (3) the substitution of the heuristic attribute in the judgment and the immediacy of the response is not rejected by the critical operations of System 2 [Kahneman and Frederick, 2002: 54].

System 2 is composed of those mental operations that are "slower, serial, effortful, more likely to be consciously monitored and deliberately controlled" [Kahneman, 2003: 698]. In general, System 2 consists of explicit cognitive processes, as opposed to mental operations that are automatic, effortless, associative and implicit. The extension of this mental process to many other heuristics is quite straightforward [Kahneman and Frederick, 2002; Kahneman, 2003].

3.1. Problem Statement Mistakes

3.1.1. Mistake-1: Not Stating the Problem in Terms of Customer Needs

Not stating the problem in terms of customer needs, but rather committing to a class of solutions causes a lack of flexibility. Identifying the true customer needs can be difficult because stakeholders often refer to both problem and solution domains—whichever comes most naturally. In systems engineering, the initial problem statement must be written before looking for solutions [Wymore, 1993].

Recommendation: Communicate with and question the customer in order to determine his or her values and needs. State the problem in terms of customer requirements [Bahill and Dean, 1999; Daniels and Bahill, 2004; Hooks and Farry, 2001; Hull, Jackson, and Dick, 2005]. Later, after gaining a better understanding of evaluation criteria and weights of importance, be open to creativity in finding alternative solutions that provide a good match to the requirements.

3.1.2. Mistake-2: Incorrect Question Phrasing

The way you phrase the question may determine the answer you get. Alluding to the problem of the formulation of public policy, Kahneman and Ritov [1994] showed their subjects (1) brief statements that looked like headlines and (2) proposed methods of intervention. Some subjects were asked to indicate their willingness to pay for the interventions by voluntary monetary contributions: other subjects were asked which intervention they would rather support.

<u>Issue M:</u>

Problem: Several Australian mammal species are nearly wiped out by hunters.

Intervention: Contribute to a fund to provide a safe breeding area for these species.

<u>Issue W:</u>

Problem: Skin cancer from sun exposure is common among farm workers.

Intervention: Support free medical checkups for threatened groups.

On being asked how much money they would be willing to contribute, most subjects indicated that they would contribute more money to provide a safe breeding area for the Australian mammal species than they would to support free medical checkups for the threatened farm workers. However, when the subjects were asked which intervention they would support, they indicated that they would rather support free medical checkups for threatened workers.

If a problem statement is vague (such as "work for the public good"), proposed solutions could vary greatly, and derive support for very different reasons and in different ways. If a problem statement is poorly written or ambiguous, dissimilar alternative solutions could remain in the solution pool, obfuscating their rational consideration, especially if the rationale for the different psychologically attractive values of the alternative solutions are not well understood [Keeney, 1992].

Recommendation: Questions designed to get a value for a criterion should be tightly coupled to the criterion.

The above example of phrasing the question is more subtle than the following one. When asked which package of ground beef they would prefer to buy, many more people chose the package labeled "80% lean," than the one labeled "20% fat."

3.1.3. Mistake-3: Substituting a Related Attribute

"Attribute substitution" occurs when a subject is assessing an attribute and substitutes a related attribute that comes more readily to mind. In effect, "people who are confronted with a difficult question sometimes answer an easier one instead" [Kahneman, 2003: 707]. When confronted with a choice among alternatives that should properly be decided by a full tradeoff study, there is a strong tendency to substitute a seemingly equivalent yet much simpler decision question in place of the tradeoff study process.

Recommendation: Sponsors of tradeoff studies should realize that a premature reduction of a tradeoff study process to a simpler decision question is a common heuristic that prevents consideration of the original multiobjective decision.

3.1.4. Mistake-4: Political Correctness

For political reasons, the top-level decision makers are often afraid to state the problem clearly and concisely. Furthermore, it is a zero-sum game. If I give more money for your research, then I have to take money away from someone else's research and politics might give their research a higher priority.

Recommendation: Outlaw political correctness.

3.2. Evaluation Criteria Mistakes

3.2.1. Mistake-1: Confounded and Dependent Criteria

Criteria may be undifferentiated or confounded. For example, in a dog-selecting tradeoff study, the criterion "Fear of Theft or Loss" confounded monetary risk from theft or loss, *and* emotional fear: The weight for "Fear of Theft or Loss" was higher than that for the monetary Cost criterion ("≤ $100/month") accumulated over a year to $1200, despite the fact that a dog typically costs less that $1200. Thus, the weights indicated that a theft or loss was expected more than once per year. Actually, the unnoted and confounded presence of an emotional fear criterion became apparent.

Recommendation-1: The determination of criteria independence must be thorough. An analyst should refer to the literature of requirements writing [Bahill and Dean, 1999; Hooks and Farry, 2001; Hull, Jackson, and Dick, 2005] to ensure completeness and differentiation of requirements, as well as proper hierarchical arrangement.

Evaluation criteria should be independent. For evaluating humans, Height and Weight are not independent: Sex (male versus female) and Intelligence Quotient are independent. In selecting a car, the following criteria are not independent: Maximum Horse Power, Peak Torque, Top Speed, Time for the Standing Quarter Mile, Engine Size (in liters), Number of Cylinders, Time to Accelerate 0 to 60 mph.

Recommendation-2: Dependent criteria should be grouped together as subcriteria. The seven subcriteria for the car given in the previous paragraph could all be grouped into the criteria Power.

3.2.2. Mistake-2: Relying on Personal Experience

"We are all prisoners of our own experience." Criteria may be chosen from the analyst's personal experience, with insufficient customer input and environmental confirmation. Uncorrelated to objective knowledge, self-assessed knowledge (what people think they know) influences their selection of search strategies [Park, Mothersbaugh, and Feick, 1994]. Aspects of sole reliance on past personal experience include preconceived notions, false memories and overconfidence in judgments [Arkes, 1981].

Recommendation: It is imperative to conduct thorough searches for objective, scientifically verifiable, knowledge.

3.2.3. Mistake-3: The Forer Effect

Previously existing criteria will be adopted if (1) the analyst believes that the criteria apply to the present problem, (2) the criteria are well presented, and (3) the analyst believes in the authority of the previous criteria writer. The analyst might fail to question or rewrite criteria from a legacy tradeoff study that originated from a perceived authority and is now seemingly adaptable to the tradeoff at hand. This is called the Forer effect. Psychologist Bertram R. Forer [1949] gave a personality test to his students. He then asked them to evaluate their personality analysis, supposedly based on their test's results. Students rated their analysis on a scale of 0 (very poor) to 5 (excellent) as to how well it applied to them. The average score was 4.26. Actually, Forer had given the same analysis to all the students. He had assembled this analysis of a generally likeable person from horoscopes. Variables that contribute to this fallacy in judgment are that the subject believes the analysis only applies to them, the subject believes in the authority of the evaluator, and the analysis lists mainly positive traits.

Recommendation: Spend time considering and formulating criteria from scratch, before consulting and possibly reusing previously written criteria. Use Value-focused Thinking [Keeney, 1992].

3.3. Weight of Importance Mistakes

3.3.1. Mistake-1: Ignoring Severity Amplifiers

When a group of people is asked to assign a weight of importance for an evaluation criterion, each person might produce a different value. Different weights arise not only from different preferences, but also from irrational severity amplifiers [Bahill and Karnavas, 2000]. These include the factors of lack of control, lack of choice, lack of trust, lack of warning, lack of understanding, manmade, newness, dreadfulness, personalization, recallability, and immediacy. Excessive disparities occur when a person assesses a weight of importance after framing the problem differently. An evaluation may depend on how the criterion affects that person, how well that person understands the alternative technologies, the dreadfulness of the results, etc. As a result, each person might assign a different weight of importance to any criterion.

Recommendation: Interpersonal variability can be reduced with education, peer review of the assigned weights, and group discussions. But be aware that people are like lemmings: if you reveal how other people are voting, then they are likely to respond with the most popular answers. It is also important to keep a broad view of the whole organization, so that criteria in one area are considered in light of all other areas. A sensitivity analysis can show how important each weight is. For unimportant weights, move on. For important weights, spend more time and money trying to get consensus: This might include showing the recommended alternatives for several different sets of weights.

3.3.2. Mistake-2: Choice Versus Calculation of Values

A weight of importance can have different values depending on whether the subject chooses the value from a predetermined set or calculates the value. For example, Tversky, Sattath, and Slovic [1988, pp. 376–377] told their subjects, "About 600 people are killed each year in Israel in traffic accidents. The ministry of transportation investigates various programs to reduce the number of casualties. Consider the following two programs, described in terms of yearly costs (in millions of dollars) and the number of casualties per year that is expected following the implementation of each program. Which program would you prefer?"

Choose a program		
	Expected casualties	Cost
Program X	500	$55M
Program Y	570	$12M

Given this "Choose a program" formulation, 67% of the subjects choose Program X, saying that "human life is priceless." However, when the same set of alternatives was setup as a calculation problem, where the subjects were asked to fill in the missing dollar value in the following box (the cell with the question marks), fewer than 4% of the subjects chose $55M or more as the worth of Program X.

Calculate a value		
	Expected casualties	Cost
Program X	500	???
Program Y	570	$12M

In this study, the subjects' mental values were different, depending on whether they choose a value from a predetermined set or made a calculation.

The weight of importance of an evaluation criterion could be determined by choice or by calculation. When a legacy tradeoff study is used to guide the present study, the weights are being determined by choice. When a tradeoff study is constructed from scratch, the weights are calculated for the current alternatives. When faced with a choice involving nonnumerical attributes, people will use principles and categories to make a decision; but when faced with a calculation problem, they will shift a great deal of attention to numbers, ratios, and differences, sometimes to the point of making narrow-minded decisions on numbers alone.

Recommendation: The use of choice and calculation should be consistent within a tradeoff study. The systems analyst might use one of the formal weight derivation tools referenced in Section 2.

3.4. Alternative Solution Mistakes

3.4.1. Mistake-1: Serial Consideration of Alternatives

When solving a problem, people often reveal a confirmation bias [Nickerson, 1998] when seizing on a hypothesis as a solution, and holding on to it until it is disproved. Once the hypothesis is disproved, they will progress to the next hypothesis and hold on to it until it is disproved [Petroski, 2003]. This confirmation bias can persist throughout a tradeoff study, as an analyst uses the whole study to try to prove that a currently favored alternative is the best.

Recommendation: Alternative solutions should be evaluated in parallel from the beginning of the tradeoff study, so that a collective and impartial consideration will permit the selection of the best alternative from a complete solution space. All alternatives should be given substantial consideration [Wickelgren, 1974]. Slovic and Fischhoff [1977] and Koriat, Lichtenstein, and Fischhoff [1980] have demonstrated the effectiveness of strategies that require equal consideration for all alternatives.

3.4.2. Mistake-2: Isolated or Juxtaposed Alternatives

Hsee et al. [1999] showed the following example. The two music dictionaries described in this box were evaluated in isolation and jointly.

Music Dictionary	Number of entries	Condition
A	10,000	Like new
B	20,000	Cover is torn

When the dictionaries were evaluated in isolation, most subjects were willing to pay more for dictionary A than for B. However, when the dictionaries were evaluated jointly, most subjects were willing to pay more for dictionary B.

The weights of importance for the number of entries and the book's condition evidentially changed. In isolation, each dictionary was judged more critically according to its condition. However, when the dictionary descriptions were juxtaposed, the number of entries became easier to compare, and the importance attached to the number of entries increased. Features that were hard to assess in separate evaluations were easier to evaluate in a comparative setting. In isolation, an assessment of the absolute values of the weights of importance was attempted; jointly, a choice among alternatives became predominant.

This phenomenon has implications for the selection of alternative solutions. Specifically, solutions that are evaluated serially (perhaps as they are conceived) may not receive the attention they would if they were evaluated in parallel with all solutions present.

Recommendation: New alternative solutions should be stored and be subject to elimination only after comparison to all alternative solutions.

3.4.3. Mistake-3: Conflicting or Dominated Criteria

Tversky and Shafir [1992] created the following experiment to analyze the effect of conflicting choices. Subjects were told, "You can either select one of these gambles or you can pay $1 to add one more gamble to the choice set. The added gamble will be selected at random from the list you reviewed."

Conflicting criteria situation			
Gamble	Winning %	Payoff	Expected value (not shown to subjects)
X	65%	$15	$9.75
Y	30%	$35	$10.40
Result: 55% of subjects chose to pay for the addition of a yet unknown gamble from a pre-reviewed set			

Next, the criteria values were changed so that one choice dominated the other and the following result occurred.

Criteria dominance situation			
Gamble	Winning %	Payoff	Expected value (not shown to subjects)
X	65%	$15	$9.75
Z	65%	$14	$9.10
Result: 30% of subjects chose to pay for the addition of a yet unknown gamble from a pre-reviewed set			

It is seen that a choice among alternatives with conflicting features will cause a continued search for alternatives, while the presence of clear dominance among the alternatives will increase the chance of the decider finalizing his or her decision.

Recommendation: Assure that the alternative solutions represent the global solution space. Do not curtail the search for alternatives when perceiving criteria dominance by one alternative.

3.4.4. Mistake-4: Failing to Consider Distinctiveness by Adding Alternatives

A similar study in medical decision making by Redelmeier and Shafir [1995] investigated distinctiveness by adding alternatives. The objective was to determine whether situations involving multiple options could paradoxically influence people to choose an option that would have been declined if fewer options were available. Surveys were mailed randomly in one of two versions to members of the Ontario College of Family Physicians, neurologists and neurosurgeons affiliated with the North American Symptomatic Carotid Endarterectomy Trial, and a group of legislators belonging to the Ontario Provincial Parliament. The following problem statement was sent to them:

> The following patients have been scheduled for carotid endarterectomy [cleaning of plaque from the arteries that supply the brain with blood], but two operating room slots have already been taken by emergency cases (more slots will not be available for 2 weeks). Of these patients, who should have a higher priority?
>
> <u>Patient M.S.</u> is a 52-year-old employed journalist with TIA's [Transient Ischemic Attack: a mini-stroke caused by temporary interruption of blood supply to an area of the brain] experienced as transient aphasia. She has had one such episode occurring ten days ago which lasted about 12 hours. Angiography shows a 70% stenosis of the left carotid. Past medical history is remarkable [noteworthy] for past alcoholism (no liver cirrhosis) and mild diabetes (diet controlled)
>
> <u>Patient A.R.</u> is a 72-year-old retired police officer with TIA's experienced as left hand paralysis. He has had two such episodes over the last three months with the last occurring one month ago. Angiography shows a 90% stenosis of the right carotid. He has no concurrent medical problems and is in generally good health.
>
> If asked for your opinion, on which patient would you operate first?
>
> One group of deciders was given just these two choices: patients M. S. and A. R: 38% chose Patient A. R.

Two patient choice	
Patient	Number of subjects choosing this patient
M. S.	109
A. R.	68
38% chose patient A. R.	

> Another group of deciders was given an additional patient.
>
> <u>Patient P.K.</u> is a 55-year-old employed bartender with TIA's experienced as transient monocular blindness. She had one such episode a week ago, that lasted less than 6 hours. Angiography shows a 75% stenosis of the ipsilateral carotid. Past medical history is remarkable for ongoing cigarette smoking (since age 15 at a rate of one pack per day).

In the group of deciders that was given all three patients, 58% chose Patient A. R.

Three patient choice	
Patient	Number of subjects choosing this patient
M. S.	57
P. K.	6
A. R.	102
58% chose patient A. R.	

Adding an additional alternative can increase decision difficulty and thus the tendency to choose a distinctive alternative. Note that the distinctiveness of the alternative was rather unnoticeable before the additional alternative was added.

Recommendation: All of the alternative solutions should be evaluated in parallel from the beginning of the tradeoff study. If an alternative must be added in the middle of a study, then the most similar alternative will lose support.

3.4.5. Mistake-5: Maintaining the Status Quo

In another similar scenario involving a patient with osteoarthritis, family physicians were less likely to prescribe a medication (53%) when deciding between two medications than when deciding about only one medication (72%). The difficulty in deciding between the two medications led some physicians to recommend not starting either.

In an experiment by Tversky and Shafir [1992], students were given \$1.50 for filling out a questionnaire. They were then offered either a metal zebra pen for their \$1.50, or they were offered a choice of a metal zebra pen or two plastic pilot pens for their \$1.50. The probability of keeping the \$1.50 was lower when they were offered only the zebra pen (25%) than when they were offered the choice of pens (53%). An increase in the conflict of the choice thus increased a decision to stay with the status quo. It seems that the students thought the increase in conflicts within the subset of pen options increased the chance of making a bad choice within the subset.

To change behavior, do not denigrate an existing concept; rather extol the virtues of a new concept, because people are more willing to accept a new concept than to reject an old one.

Recommendation: The more alternatives that exist and the more complicated the decision, the more the status quo will be favored. Do not needlessly increase the number of alternatives in a tradeoff study. More alternatives increase the difficulty of the decision. However, in the very beginning of a project it is good to have many alternatives in order to better understand the problem and the requirements.

3.5. Evaluation Data Mistakes

3.5.1. Mistake-1: Relying on Personal Experience

Guesses for evaluation data may faultily come from personal memory. People may be oblivious to things they have not experienced, or they may think that their limited experience is complete. What people think they know may be different from what they actually know [Schacter, 1983].

Recommendation: The source of evaluation data for public projects must be subject to peer and public review. Decision analysts must be willing to yield absolute control over choosing evaluation data.

3.5.2. Mistake-2: Failing to Consider Both Magnitude and Reliability

People tend to judge the validity of data first on its magnitude or salience ("strength"), and then according to its reliability ("weight") [Griffin and Tversky, 1992]. Therefore, data with outstanding magnitudes but poor reliability are likely to be chosen and used.

Recommendation: Either data with uniform reliability should be used, or the speciousness of data should be taken into account in the Risk portion of a tradeoff study.

3.5.3. Mistake-3: Judging Probabilities Poorly

Probabilistic evaluation data usually comes from blue-sky guesses by domain experts. Later, probabilistic guesses are refined with quantitative estimates, and then with simulations and finally prototypes. So, data often come from human evaluation, but humans are terrible at estimating probabilities, especially for events with probabilities near 0 and 1 [Tversky and Kahneman, 1992]. In general, a refined understanding of probability theory is not usual [Gigerenzer, 1991].

Recommendation: Distrust human probability estimates, especially for improbable and very probable events

3.5.4. Mistake-4: Ignoring the First Measurement

Often when a measurement (test) reveals an unexpected result, the physician and/or the patient will ask for a second measurement. If the second measurement is pleasing, then the first measurement is discarded and only the result of the last measurement is recorded. This is probably related to confirmation bias: the decider rejects the results of all tests until one confirms the decider's preconceived notions.

Recommendation: If there is no evidence showing why the first measurement was in error, then it should not be discarded. A reasonable strategy would be to record the average of the two measurements. For example, if you take your blood pressure and the result is abnormally high, then you might measure it again. If the second measurement indicates that blood pressure is near the normal range, and you do not have proof that the first reading was a mistake, then do not record only the second reading, either record both measurements or the average of the two readings.

3.5.5. Mistake-5 Anchoring

When estimating numerical values, a person's first impression dominates all further thought. In this example from Piattelli-Palmarini [1994] people were shown a wheel of fortune with numbers from 1 to 100. The wheel was spun and then the subjects were asked to estimate the number of African nations in the United Nations. If the wheel showed a small number, like 12, the subjects inevitably underestimated the correct number. If the number on the wheel were large, like 92, the subjects overestimated the correct number.

Recommendation: View the problem from different perspectives. Use different starting points.

3.5.6. Mistake-6 Anchoring and the Status Quo

Normally, you fill out a tradeoff study matrix row by row and the status quo is the alternative in the first column. Therefore, the values of the status quo are the anchors for estimating the other data. Unfortunately, the status quo is likely to have low values for performance and high values for cost, schedule, and risk. But at least the anchoring alternative is known, is consistent, and you have control over it.

Recommendations: Make the status quo the alternative in the first column. In one iteration, evaluate the scores left to right and in the next iteration evaluate them right to left.

3.6. Scoring Function Mistakes

3.6.1. Mistake-1: Treating Gains and Losses Equally

People do not Treat Gains and Losses Equally. People prefer to avoid losses more than to acquire gains. Prospect Theory [Kahneman and Tversky, 1979] suggests that psychologically losses are twice as powerful as gains. Would you rather get a 5% discount, or avoid a 5% penalty? Most people would rather avoid the penalty. In a tradeoff study, you will get a different result if the scoring function expresses losses rather than gains.

The **Pinewood Derby Tradeoff Study** is a real-world tradeoff study that also serves as a reference model. The original study was published in Chapman, Bahill, and Wymore [1992, Chapter 5]. Pinewood [2006] was subsequently implemented in Excel with a complete sensitivity analysis.

> Over 80 million people have participated in Cub Scout Pinewood Derbies. Pinewood is a case study of the design of a Cub Scout Pinewood Derby for one particular scout pack. The system helps manage the entire race from initial entry through final results. Many alternatives for race format, scoring, and judging are presented [Chapman, Bahill and Wymore, 1992: 83].

The Pinewood Derby tradeoff study had the following criteria:

Percent Happy Scouts
Number of Irate Parents

Because people evaluate losses and gains differently, the preferred alternatives might have been different if they had used:

Percent Unhappy Scouts
Number of Ecstatic Parents

When we showed people Figure 1 and asked, "How would you feel about an alternative that gave 90% happy scouts?" they typically said, "It's pretty good." In contrast, when we showed people Figure 2 and asked, "How would you feel about an alternative that gave 10% unhappy scouts?" they typically said, "It's not very good." When we allowed them to change the parameters, they typically pushed the baseline for the Percent Unhappy Scouts (Fig. 2) scoring function to the left.

Human unequal treatment of gains and losses suggests that scoring functions in a tradeoff study should uniformly express either gains or losses. Principles of linguistic comprehensibility suggest that criteria should always be phrased in a positive manner, for example, use Uptime rather than Downtime, use Mean Time Between Failures rather than Failure Rate, use Probability of Success, rather than Probability of Failure. Finally, when using scoring functions make sure more output is better.

In a less subtle experiment, when subjects were asked whether they would approve surgery to themselves in a hypothetical medical emergency, many more people accepted surgery when the chance of survival was given as 80% than when the chance of death was

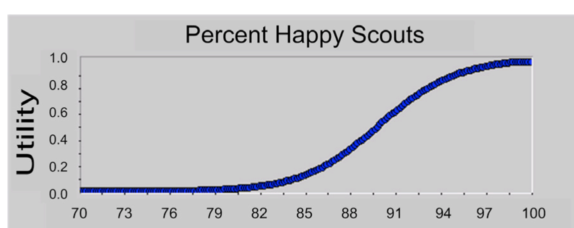


Figure 1. Scoring function for percent happy scouts from the Pinewood Derby tradeoff study.

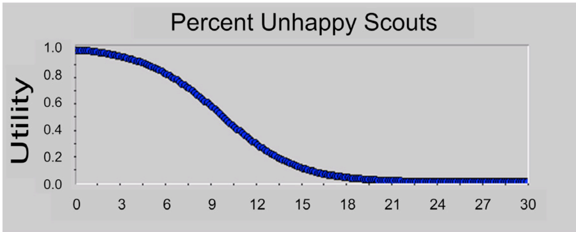


Figure 2. Scoring function for percent unhappy scouts.

given as 20%. Also see Lichtenstein and Slovic [1971] for examples of preference reversals.

Recommendation: Scoring functions in a tradeoff study should uniformly express gains rather than losses.

3.6.2. Mistake-2: Not Using Scoring Functions

In "infrastructure bias," the location and availability of preexisting infrastructure such as roads and telecommunication facilities influences future economic and social development. In science, an availability bias is at work when existing scientific work influences future scientific observations. For example, when sampling pollutants, most samples may be taken in towns or near roads, as they are the easiest places to get to. Other examples occur in astronomy and particle physics, where the availability of particular kinds of telescopes or particle accelerators acts as constraints on the types of experiments performed.

Most tradeoff studies that we have observed in industry did not use scoring functions. In some cases, scoring functions were explained in the company engineering process, but they were not convenient, hence they were not used.

Recommendation: The four Wymorian standard scoring functions [Wymore, 1993] of Figure 3 (or similar functions, or fuzzy sets or utility functions) should be used in tradeoff studies. Easy-to-use scoring functions, such as those located at http://www.sie.arizona.edu/sysengr/slides/SSF.zip, should be referenced in company systems engineering processes.

3.6.3. Mistake-3 Anchoring and Scoring Functions

When estimating numerical values, a person's first impression dominates all further thought.

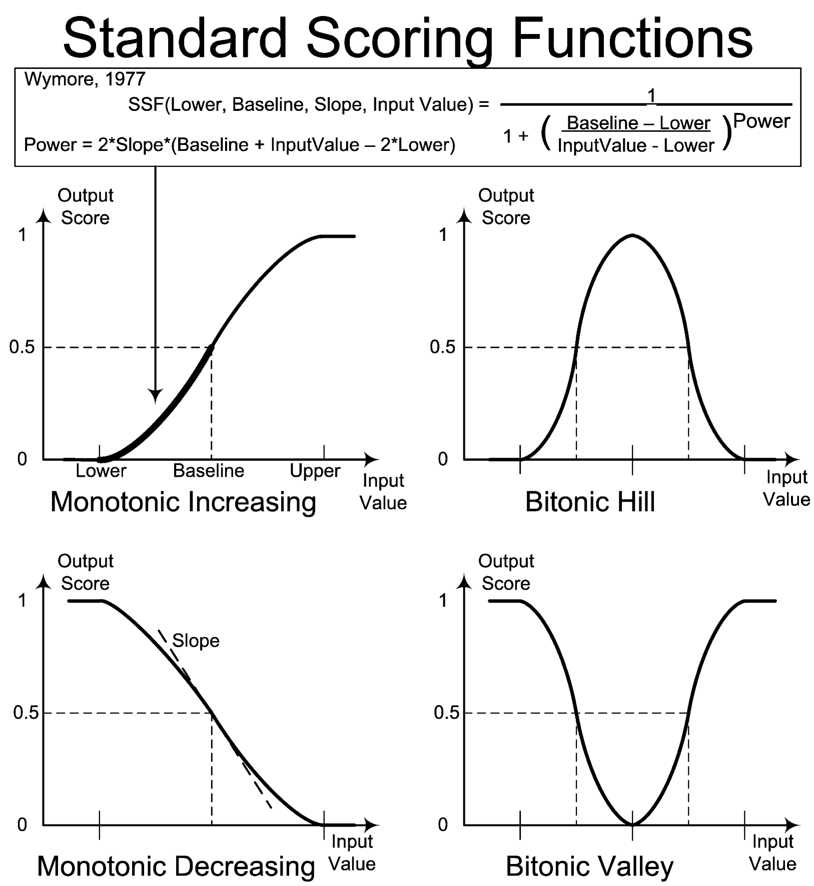


Figure 3. The basic Wymorian scoring functions.

Recommendation: View the problem from different perspectives. Use different starting points. When estimating values for parameters of scoring functions, think about the whole range of expected values for the parameters. For example, when estimating the baseline for the percent Happy Scouts of Figure 1, the systems engineer could show the expert two scoring functions one with a baseline of 80 and another with a baseline of 98. Then allow the expert to express his or her preference.

3.7. Score Mistakes

When scores are derived directly from customer preferences, they may be subject to cognitive biases of the customer. However, when they are derived from evaluation data and scoring (utility) functions that are correct, then the scores should be free of cognitive errors, because they are computed without human intervention. However, there is one frequent mistake in computing scores.

3.7.1. Mistake-1: Implying False Precision

The most common mistake that we have seen in tradeoff studies is false precision. For example, a tradeoff analyst might ask a subject matter expert to estimate values for two criteria. The expert might say something like, "The first criterion is about 2 and the second is around 3." The analyst puts these numbers into a calculator and computes the ratio as 0.666666667. This is nonsense, but these nine significant digits may be used throughout the tradeoff study. The Forer Effect might explain this. The analyst believes that the calculator is an impeccable authority in calculating numbers. Therefore, what the calculator says must be true.

Recommendation: Use significant digit methodology. Furthermore, in numerical tables, print only a sufficient number of digits after the decimal place as is necessary to show a difference between the preferred alternatives.

3.8. Combining Function Mistakes

3.8.1. Mistake-1: Lack of Knowledge

The average engineer is not familiar with the nuances of combining functions and their behavior specific to tradeoff studies. This is a judgment derived from Bahill's 20-year search for good tradeoff studies in industry.

Recommendation: Training with combining functions is necessary. Discussions of combining functions are found in Keeney and Raiffa [1976], Daniels, Werner, and Bahill [2001], and Smith [2006].

3.8.2. Mistake-2: Lack of Availability

Software is equipped with limited types of combining functions. For example, one of the best commercial tools, Expert Choice, has only the Sum and the Product combining functions. Most others have only the Sum.

Recommendation: Spreadsheet-formulated tradeoff studies have the greatest potential for combining function variety.

3.9. Preferred Alternative Mistakes

3.9.1. Mistake-1: Overconfidence in Subjective Choice

Tradeoff studies are often started with overconfidence. Then the analyst prefers to maintain a state of overconfidence without examining details. Griffin and Tversky [1992: 411] state:

> People are often more confident in their judgments than is warranted by the facts. Overconfidence is not limited to lay judgment or laboratory experiments. The well-publicized observation that more than two-thirds of small businesses fail within 4 years [Bradstreet, 1967] suggests that many entrepreneurs overestimate their probability of success [Cooper, Woo, and Dunkelberg, 1988].

Recommendation: For this bias, there is no better teacher than performing tradeoff studies and then presenting the results at reviews that demand high-quality work in all tradeoff study components.

3.9.2. Mistake-2: Obviating Expert Opinion

The analyst holds a circular belief that expert opinion or review is not necessary because no evidence for the need of expert opinion is present. This is especially true if no expert has ever been asked to comment on the tradeoff study.

All humans store about seven units or "chunks" of information at a time [Miller, 1956], in short-term memory, irrespective of skill level. However, chess masters' chunks are larger and richer in information than amateurs' chunks [Chase and Simon, 1973]. Thus, a novice often "can't see the forest for the trees," and cannot perceive the refinement in an expert's thought.

Recommendation: Experts should be sought formally or informally to evaluate tradeoff study work.

3.10. Sensitivity Analysis Mistakes

3.10.1. Mistake-1: Lack of Training

Most personnel are not well trained in the machinery and methods of sensitivity analysis. They often fail to compute second- and higher-order partial derivatives. When estimating partial derivatives, they often use too large a step size. When estimating partial derivatives of functions of two parameters they often use the wrong formula; they use a formula with two instead of four numerator terms. A recent paper by Smith et al. [2005] has shown that interactions among parameters in tradeoff studies can be very important, step sizes for the approximation of effects should be very small, and second-order derivatives should be calculated accurately. It is expected that only the best-trained personnel will know of such results, illustrating the gap between truth and training.

Recommendation: Investments in sensitivity analysis training must be made. Perhaps enabling software can substitute for much sensitivity analysis knowledge. Karnavas, Sanchez, and Bahill [1993], Saltelli et al. [2004], and Cacuci [2005] describe the use of sensitivity analyses.

4. DISCUSSION

The mental mistakes presented in this paper represent the principal errors in judgment to which a tradeoff analyst will most likely fall prey. Recommendations for dealing with these mental mistakes have been given.

Humans usually consider alternatives in series, and often hastily choose one alternative after having their attention anchored, or fixated, to only one or a few criteria. Also, humans tend to choose one alternative based on conclusions derived from their favorite theories. Ideally, decision analysts should consider the complete decision space, or a complete set of alternatives, and then use a shrinking set of hypotheses brought about by conclusions based on experimentation, empirical data, and data analysis. In other words, we recommend that, for appropriate problems, the decision-maker should eschew typical human serial processing of alternatives and instead evaluate the alternatives in parallel in a tradeoff study.

In order to choose rationally among alternatives, the decision analyst should be aware of mental mistakes, and employ good debiasing recommendations and techniques. Advice from subject matter experts should augment single-handed judgment. Decision analysts should also have a complete understanding of the mathematical methods that allow the parallelization of human decision processes through tradeoff studies, and be able to apply them without error.

Another conclusion is that newly presented, complex choices among alternatives should not be simplified into feeling-based "gut decisions." Nonexperts should realize that making important decisions with what is perceived to be a mentally-holistic, feeling-based approach, without regard to the needed rationality, is risky. In order to establish rationality, all the components of the decision must be carefully considered until clarity as to their arrangement and magnitude results. This is possible by focusing on the elements individually and then collectively. The higher-level decision then becomes a calculation based on a broad base of rationally considered components.

There is significant interest among decision makers about complex decisions made in the instant of perception [Gladwell, 2005]. It should be noted that experts capable of making such judgments have spent long periods of time in training [Klein, 1998], during which they have individually, sequentially, repeatedly, comparatively, and rationally examined all the components of the decision. Any preconscious parallelization occurring in such an expert's brain is reproduced in the parallel structure of a tradeoff study, which is ultimately based on hard data analysis.

Limitations. Limited time and resources guarantee that a tradeoff study will never contain all possible criteria. Therefore, tradeoff studies never produce *optimal* solutions: They produce *satisficing* solutions [Simon, 1955, 1957]. A tradeoff study reflects only one view of the problem. Different tradeoff analysts might choose different criteria and weights and therefore paint a different picture of the problem. In this paper, we have ignored human decision-making mental mistakes for which we have no suggested corrective action, such as closed mindedness, lies, conflict of interest, and favoritism.

In a tradeoff study, it is natural to ask, "Have we written enough criteria? Have we studied enough alternatives?" One way to assess the completeness of a modeling study is to complete a Zachman framework for the study [Bahill, Botta, and Daniels, 2006].

We have studied three tradeoff studies that had variability or uncertainty statistics associated with each evaluation datum. Two of these were personal observations of Bahill and one was published [Ullman and Spiegel, 2006]. In these studies, the uncertainty statistics were carried throughout the whole computational process, so that at the end the recommended alternatives had associated uncertainty statistics. However, these studies used approaches too complicated for most practical purposes. Therefore, in our tradeoff studies we do not try to accommodate uncertainty, changes in uncertainty, or dependencies in the evaluation data. It can be done, but we warn that it is complicated.

Good industry practices for ensuring the success of tradeoff studies include having teams evaluate the data, evaluating the data in many iterations and expert review of the results. It is important that the expert review teams have reviewers that are external to the project and that the reviewers consider design decisions as well as simply checking to ensure that required tasks were done. Reviews are often hampered by failure to allow external reviewers access to proprietary or classified data [Leveson, 2007].

It is important to remember that the output of a tradeoff study is only recommendations. These recommendations must then be blended with non-quantitative assessments of organizational vision, goals, culture and values.

5. SUMMARY EXAMPLE

Invincibility bias. Many bad decisions can be attributed to the decision-maker's sense of invincibility. Teen-age boys are notorious for thinking, "I won't get caught: I can't get hurt: I will avoid car accidents: I won't cause an unwanted pregnancy: I won't get sexually transmitted disease:" and that, "They can't do that to me." Many other people think, "Nobody is going to steal my identity: I'll be able to quit any time I want to: I don't need sun screen, I won't get skin cancer:" and "I don't have to back up my hard drive, my computer won't crash." The Spanish Armada was thought to be invincible in 1588. The Japanese thought they were invincible at Pearl Harbor in 1941. The German military thought it was invincible as it stormed across Europe in WW II. And, of course, in 1912, the White Star line said that the Titanic was "unsinkable." The invincibility bias will affect the risk analysis, the problem statement, and the selection of criteria for the tradeoff study.

The original design for the **RMS Titanic** called for 64 lifeboats, but this number was reduced to 20 before its maiden voyage: This was a tradeoff mistake. The Chief Designer wanted 64 lifeboats, but the Program Manager reduced it to 20 after his advisors told him that only 16 were required by law. The Chief Designer resigned over this decision. The British Board of Trade regulations of 1894 specified lifeboat capacity: for ships over 10,000 tons, this lifeboat capacity was specified by volume (5,500 cubic feet), which could be converted into passenger seats (about 1000) or the number of lifeboats (about 16). So, even though the Titanic displaced 46,000 tons and was certified to carry 3,500 passengers, its 20 lifeboats complied with the regulations of the time. But let us go back to the design decision to reduce the number of lifeboats from 64 to 20. What if they had performed the following hypothetical tradeoff study? In Table I, the weights of importance range from 0 to 10, with 10 being the most important and the evaluation data (scores) also range from 0 to 10, with 10 being the best. For simplicity, we have not used

Table I. An Apocryphal Tradeoff Study for the RMS Titanic

	Weights of importance		Alternatives and their evaluation data			
Criteria	Program Manager's Weights	Chief Designer's Weights	10 lifeboats	20 lifeboats	30 lifeboats	64 lifeboats
Will it satisfy the Board of Trade regulations? (yes, no)	10	10	0	10	10	10
Amount of deck space required for the lifeboats (ft^2)	2	2	10	8	4	2
Possible perception that the ship is unsafe caused by the presence of a large number of lifeboats	6	2	10	8	4	2
Cost to purchase, install, maintain, launch and operate the lifeboats (£)	9	4	10	8	4	2
Percentage of passengers and crew that could be accommodated in lifeboats, if all lifeboats were launched full of people	1	10	2	4	6	10
Alternative ratings produced by summating the Program Manager's weights times scores			172	**240**	174	144
Alternative ratings produced by summating the Chief Designer's weights times scores			100	204	192	**216**

scoring functions, so the evaluation data are also the scores.

The magnitudes of the Program Manager's and the Chief Designer's alternative ratings are not in themselves important. What is important is that they indicate different preferred alternatives, which result from different sets of weights of importance.

The Program Manager might have had **overconfidence in his subjective choice** of 20 lifeboats**.** If he had done this tradeoff study, might the Program Manager have rethought his decision to use only 20 lifeboats?

Many bad decisions can be attributed to the decision-maker's sense of **invincibility**. In 1912, the White Star line said that the Titanic was "unsinkable." If the Program Manager had not also believed this, would he have authorized more lifeboats?

If the Program Manager understood the **Forer effect** (that an analyst might fail to question or rewrite criteria that originated from a perceived authority), might he have reassessed the Board of Trade's regulation for 16 lifeboats?

The Program Manager and the Chief Designer did not do a tradeoff study. They merely discussed the 20- and 64-lifeboat alternatives. If they had understood **distinctiveness by adding alternatives** and had done this tradeoff study with the addition of the 10- and 30-lifeboat alternatives, is it likely that the Program Manager would have chosen a different alternative?

The mandatory requirement of satisfying the Board of Trade regulations ruled out the 10-lifeboat alternative. A **sensitivity analysis** of the remaining tradeoff study shows that the most important parameter from the Program Manger's perspective is the weight of importance for the Cost criterion and that the most important parameter from the Chief Designer's perspective is the weight of importance for the "Percentage of passengers and crew that could be accommodated in lifeboats" criterion. Therefore, the Program Manager should have spent more time assessing the **magnitude and reliability** of the values of these parameters. In fact, he should have noted the importance of safety regulations, and questioned whether the Board of Trade's regulation for 16 lifeboats was reliable for the new, larger Titanic design.

6. CONCLUSION

Humans usually consider alternatives in series, and often hastily choose one alternative after having their attention anchored to only one or a few criteria. In order to make good, rational choices among alternatives, a decision-maker's awareness of cognitive biases must increase. Limited awareness and possible consequential irrationality can cause erroneous estimation or misuse of tradeoff study components. Tradeoff study components should be examined individually. The higher-level decision then becomes a calculation resting on a broad base of rationally considered assumptions, components, and calculation methods. Decision-makers should understand the mathematical methods that allow the parallelization of human decision processes through tradeoff studies.

REFERENCES

G. Ainslie, Breakdown of will, Cambridge University Press, New York, 2001.

H.R. Arkes, Impediments to accurate clinical judgment and possible ways to minimize their impact, J Consult Social Psychol 49 (1981), 323–330.

A.T. Bahill, Mathematical summary of multicriterion decision-making methods, p. 24, http:/www.sie.arizona.edu/sysengr/slides/tradeoffMath.doc, retrieved May 2007.

A.T. Bahill and W. L. Chapman, A tutorial on quality function deployment, Engr Man J, 5(3) (1993) 24–35.

A.T. Bahill and C. Briggs, The systems engineering started in the middle process: A consensus of system engineers and project managers, Syst Eng, 4(2) (2001) 156–167.

A.T. Bahill and F.F. Dean, "Discovering system requirements," Handbook of systems engineering and management, A.P. Sage and W.B. Rouse (Editors), Wiley, New York, 1999, pp. 175–220.

A.T. Bahill and W.J. Karnavas, Risk analysis of a pinewood derby: A case study, Syst Eng 3(3) (2000), 143–155.

A.T. Bahill, R. Botta, and J. Daniels, The Zachman framework populated with baseball models, J Enterprise Architecture 2(4) (2006), 50–68.

L.R. Beach and T. Connolly, The psychology of decision making: People in organizations, Sage, Thousand Oaks, CA, 2005.

D. Bradstreet, Patterns of success in managing a business, Dun & Bradstreet, New York, 1967.

D.M. Buede, The engineering design of systems: Models and methods, Wiley, New York, 2000.

D.G. Cacuci, Sensitivity and uncertainty analysis, Chapman & Hall/CRC Press, Boca Raton, FL, 2005.

W.L. Chapman, A.T. Bahill, and A.W. Wymore, Engineering modeling and design, CRC Press, Boca Raton, FL, 1992.

W.G. Chase and H.A. Simon, Perception in chess, Cognitive Psychol 4 (1973), 55–81.

CMMI, Capability Maturity Model Integration, Software Engineering Institute, Pittsburgh, PA, 2006, http://www.sei.cmu.edu/cmmi/, retrieved March 2006.

A.C. Cooper, C.Y. Woo, and W.C. Dunkelberg, Entrepreneurs' perceived chances for success, J Bus Venturing 3 (1988), 97–108.

J. Daniels and A.T. Bahill, The hybrid process that combines traditional requirements and use cases, Syst Eng 7(4) (2004), 303–319.

J. Daniels, P.W. Werner, and A.T. Bahill, Quantitative methods for tradeoff analyses, Syst Eng 4(3) (2001), 190–212.

DAR, DAR basics: Applying decision analysis and resolution in the real world, Software Engineering Institute, Pittsburgh, PA, 2004, http://www.sei.cmu.edu/cmmi/presentations/sepg04.presentations/dar.pdf, retrieved March 2006.

W. Edwards, How to use multiattribute utility analysis for social decision making, IEEE Trans Syst Man Cybernet SMC-7 (1977), 326–340.

B.R. Forer, The fallacy of personal validation: A classroom demonstration of gullibility, J Abnormal Psychol 44 (1949), 118–121.

G. Gigerenzer, How to make cognitive illusions disappear: Beyond “heuristics and biases,” Eur Rev Social Psychol 2(4) (1991), 83–115.

M. Gladwell, Blink: The power of thinking without thinking, Little Brown, New York, 2005.

D. Griffin and A. Tversky, The weighting of evidence and the determinants of confidence, Cognitive Psychol 24 (1992), 411–435.

Y.Y. Haimes, Risk modeling, assessment, and management, Wiley, Hoboken, NJ, 2004.

J.S. Hammond, R.L. Keeney, and H. Raiffa, The hidden traps in decision making, HBR Classic, Harvard Bus Rev 184(1) (2006), 118–132.

I.F. Hooks and K.A. Farry, Customer-centered products: Creating successful products through smart requirements management, AMACOM, New York, 2001.

C.K. Hsee, G. Loewenstein, S. Blount, and M. Bazerman, Preference reversals between joint and separate evaluations of options: A theoretical analysis, Psychol Bull 125(5) (1999), 576–590.

E. Hull, K. Jackson, and J. Dick, Requirements engineering, Springer, London, 2005.

INCOSE, INCOSE systems engineering handbook v2a, Seattle, 2004, Xhttp://www.incose.org/ProductsPubs/incosestore.aspx, retrieved March 2006.

D. Kahneman, A perspective on judgment and choice: Mapping bounded rationality, Amer Psychol 58(9) (2003), 697–720.

D. Kahneman and S. Frederick, “Representativeness revisited: Attribute substitution in intuitive judgment,” Heuristics and biases, T. Gilovich, D. Griffin, and D. Kahneman (Editors), Cambridge University Press, New York, 2002, pp. 49–81.

D. Kahneman and I. Ritov, Determinants of stated willingness to pay for public goods: A study in the headline method, J Risk Uncertainty 9 (1994), 5–38.

D. Kahneman and A. Tversky, Prospect theory: An analysis of decision under risk, Econometrica 46(2) (1979), 171–185.

D. Kahneman, P. Slovic, and A. Tversky (Editors), Judgment under uncertainty: Heuristics and Biases, Cambridge University Press, Cambridge, MA, 1982.

W.J. Karnavas, P.J. Sanchez, and A.T. Bahill, Sensitivity analyses of continuous and discrete systems in the time and frequency domains, IEEE Trans Syst Man Cybernet SMC-23 (1993), 488–501.

R.L. Keeney, Value-focused thinking: A path to creative decisionmaking, Harvard University Press, Cambridge, MA, 1992.

R. L. Keeney, Common mistakes in making value trade-offs, Oper Res 50(6) (2002), 935–945.

R.L. Keeney and H. Raiffa, Decisions with multiple objectives: Preferences and value tradeoffs, Wiley, New York, 1976.

C.W. Kirkwood, “Decision analysis,” Handbook of systems engineering and management, A.P. Sage and W.B. Rouse (Editors), Wiley, Hoboken, NJ, 1999, pp. 1119–1145.

G. Klein, Sources of power: How people make decisions, MIT Press, Cambridge, MA, 1998.

R.M. Koriat, S. Lichtenstein, and B. Fischhoff, Reasons for confidence, J Exper Psychol Hum Learn Memory 6 (1980), 107–118.

N.G. Leveson, System safety engineering: back to the future, http://sunnyday.mit.edu/book2.html, retrieved May 2007.

S. Lichtenstein, and P. Slovic, Reversals of preferences between bids and choices in gambling decisions, J Exper Psychol 89 (1971), 46–55.

K.R. MacCrimmon, “An overview of multiple objective decision making,” Multiple criteria decision making, J.L. Cochrane, and M. Zelany (Editors), University of South Carolina Press, Columbia, SC, 1973, pp. 18–43.

B. Martino De, D. Kumaran, B. Seymour, and R.J. Dolan, Frames, biases, and rational decision-making in the human brain, Science 313(5787) (2006), 684–687.

G.A. Miller, The magic number seven, plus or minus two: Some limits on our capacity for processing information, Psychol Rev 63 (1956), 81–87.

R.S. Nickerson, Confirmation bias: A ubiquitous phenomenon in many guises, Rev Gen Psychol 2 (1998), 175–220.

C.W. Park, D.L. Mothersbaugh, and L. Feick, Consumer knowledge assessment, J Consumer Res 21 (1994), 71–82.

H. Petroski, Framing hypotheses: A cautionary tale, Amer Scientist 91(2) (2003), 18–22.

M. Piattelli-Palmarini, Inevitable illusions: How mistakes of reason rule our minds, Wiley, New York, 1994.

Pinewood, Pinewood Derby tradeoff study, Tucson, AZ, 2006, http://www.sie.arizona.edu/sysengr/pinewood/pinewood.pdf.

D. Redelmeier and E. Shafir, Medical decision making in situations that offer multiple alternatives, JAMA 273(4) (1995), 302–305.

RSAS,"Press release: Nobel Prize in Economics 2002, The Royal Swedish Academy of Sciences, Stockholm, 2002, http://nobelprize.org/nobel_prizes/economics/laureates/2002/press.html, retrieved September 2002.

T. Saaty, The analytic hierarchy process: Planning, priority setting, resource allocation, McGraw-Hill, New York, 1980.

A.P. Sage (Editor), Concise encyclopedia of information processing in systems & organizations, Pergamon Press, Oxford, UK, 1990.

A. Saltelli, S. Tarantola, F. Campolongo, and M. Ratto, Sensitivity analysis in practice: A guide to assessing scientific models, Wiley, Hoboken, NJ, 2004.

D.L. Schacter, Feeling of knowledge in episodic memory, J Exper Psychol Learn Memory Cognition 9 (January 1983), 39–54.

E. Shafir (Editor), Preferences, belief and similarity: Selected writings, The MIT Press, Cambridge, MA, 2004.

H.A. Simon, A behavioral model of rational choice, Quart J Econ 59 (1955), 99–118.

H.A. Simon, Models of man: Social and rational, Wiley, New York, 1957.

P. Slovic and B. Fischhoff, On the psychology of experimental surprises, J Exper Psychol Hum Perceptions Performance 3 (1977), 544–551.

E.D. Smith, Tradeoff studies and cognitive biases, Ph.D. dissertation, Systems and Industrial Engineering, University of Arizona, Tucson, AZ, 2006.

E.D. Smith, F. Szidarovszky, W.J. Karnavas, and A.T. Bahill, Sensitivity analysis, a powerful system validation tool, University of Arizona, Tucson, AZ, 2005.

F. Strack, L.L. Martin, and N. Schwarz, Priming and communication: The social determinants of information use in judgements of life-satisfaction, Eur J Social Psychol 18 (1988), 429–442.

A. Tversky and D. Kahneman, Judgment under uncertainty: Heuristics and biases, Science 185 (1974), 1124–1131.

A. Tversky and D. Kahneman, The framing of decisions and the psychology of choice, Science 211 (1981), 453–458.

A. Tversky and D. Kahneman, Advances in prospect theory: Cumulative representation of uncertainty, J Risk Uncertainty 5 (1992), 297–323.

A. Tversky and D. Kahneman, "Advances in prospect theory: Cumulative representation of uncertainty," in Choices, Values and Frames, A. Tversky and D. Kahneman (Editors), Cambridge University Press / Russell Sage Foundation, New York and Cambridge UK, 2000, pp 44–65.

A. Tversky and E. Shafir, The disjunction effect in choice under uncertainty, Psychol Sci 3(5) (1992), 305–309.

A. Tversky, S. Sattath, and P. Slovic, Contingent weighting in judgment and choice, Psychol Rev 95(3) (1988), 371–384.

D.G. Ullman and B. Spiegel, Trade studies with uncertain information, 16th Annu Int Symp INCOSE, paper 9.3.1 CD ROM, 2006.

M. Weber and K. Borcherding, Behavioral influences on weight judgments in multiattribute decision making, Eur J Oper Res 26 (1993), 35–41.

W.A. Wickelgren, How to solve problems: Elements of a theory of problems and problem solving, Freeman, San Francisco, 1974.

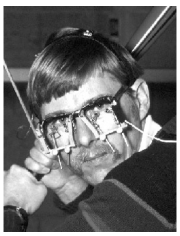
A.W. Wymore, Model-based systems engineering, CRC Press, Boca Raton, FL, 1993.

A.W. Wymore, The nature of research in systems engineering, USC Conf Syst Eng Res, April 15–16, 2004, Los Angeles, CA, paper 119, CD ROM, 2004.

Eric Smith earned a B.S. in Physics in 1994, an M.S. in Systems Engineering in 2003, and his Ph.D. in Systems and Industrial Engineering in 2006 from the University of Arizona, Tucson. He is currently a Visiting Assistant Professor in the Department of Engineering Management and Systems Engineering, at the University of Missouri at Rolla.

Young Jun Son is an Associate Professor of Systems and Industrial Engineering at the University of Arizona in Tucson. He received his Ph.D. in Industrial and Manufacturing Engineering from the Pennsylvania State University, University Park, in 2000. He is an associate editor of the *International Journal of Modeling and Simulation* and the *International Journal of Simulation and Process Modeling*. He is a recipient of the Society of Manufacturing Engineers 2004 M. Eugene Merchant Outstanding Young Manufacturing Engineer Award, the Institute of Industrial Engineers 2005 Outstanding Young Industrial Engineer Award, and the Industrial Engineering Research Conference 2005 Best Paper Award of the Modeling and simulation Track. Son has authored or co-authored over 60 publications in refereed journals and conferences, primarily on developing the new field of Distributed Federation of Multi-Paradigm Simulations and expanding the field of Computer Integrated Control to encompass the Extended Manufacturing Enterprise.

Massimo Piattelli-Palmarini is a Professor in the Cognitive Science Program, the Department of Psychology and the Department of Management and Policy at the University of Arizona. He obtained his doctorate in Physics at the University of Rome in 1968 and was principal research scientist at the Cognitive Science Center of MIT (Cambridge, MA) from 1985 to 1994.



Terry Bahill is a Professor of Systems Engineering at the University of Arizona in Tucson. He received his Ph.D. in Electrical Engineering and Computer Science from the University of California, Berkeley, in 1975. Bahill has worked with BAE Systems in San Diego, Hughes Missile Systems in Tucson, Sandia Laboratories in Albuquerque, Lockheed Martin Tactical Defense Systems in Eagan, MN, Boeing Information, Space and Defense Systems in Kent, WA, Idaho National Engineering and Environmental Laboratory in Idaho Falls, and Raytheon Missile Systems in Tucson. For these companies he presented seminars on Systems Engineering, worked on system development teams, and helped them describe their Systems Engineering Process. He holds a U.S. patent for the Bat Chooser, a system that computes the Ideal Bat Weight for individual baseball and softball batters. He received the Sandia National Laboratories Gold President's Quality Award. He is a Fellow of the Institute of Electrical and Electronics Engineers (IEEE), of Raytheon and of the International Council on Systems Engineering (INCOSE). He is the Founding Chair Emeritus of the INCOSE Fellows Selection Committee. His picture is in the Baseball Hall of Fame's exhibition "Baseball as America." You can view this picture at http://www.sie.arizona.edu/sysengr/.